Technical Research Note 216

AD

AD707375

COMPARISON OF COMPUTER-SIMULATED CONVENTIONAL AND BRANCHING TESTS

Carrie W. Waters

MILITARY SELECTION RESEARCH DIVISION

BESRL

Behavior and Systems Research Laboratory
Office of the Chief of Research and Development
U. S. Army

March 1970

41

Technical Research Note 216

AD

COMPARISON OF COMPUTER-SIMULATED CONVENTIONAL AND BRANCHING TESTS

Carrie W. Waters

A. G. Bayroff, Senior Task Leader

MILITARY SELECTION RESEARCH DIVISION
Edmund F. Fuchs, Chief

BEHAVIOR AND SYSTEMS RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

Room 239, The Commonwealth Building
1320 Wilson Boulevard, Ar ington, Virginia 22209

March 1970

Army Project Number
2Q024701A721

Enlisted Manpower

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The ENLISTED MANPOWER Work Unit conducts a continuing research program to maintain and improve techniques and procedures for screening potential enlisted manpower. Objectives are 1) to develop new forms of screening measures for use by the Army and the other services so as to assess more effectively the trainability and usability of potential enlisted personnel; 2) to develop new reference measures for use as standards in developing screening and classification tests for all the services; and 3) to improve methods for extracting predictive information from screening tests.

As one avenue to development of technical information that can contribute to more effective input screening, the feasibility of programmed testing is being investigated. BESRL has conducted several experimental and theoretical studies of branching tests in which testing is individualized by having test questions so programmed that an examinee who answers a test item correctly is presented next with a more difficult item and an examinee who answers incorrectly is presented with an easier item. By contrast, in conventional tests all examinees answer the same items presented in the same order. The present publication reports on a comparison of a variety of computer-simulated conventional and branching tests.

The entire ENLISTED MANPOWER Work Unit is responsive to special requirements of the Deputy Chief of Staff for Personnel, as well as to requirements of RDT&E Project 2Q024701A721, "Selection and Behavioral Evaluation," FY 1970 Work Program.

J. E. Uhlaner

J. E. UHLANER, Director
Behavior and Systems
Research Laboratory

COMPARISON OF COMPUTER-SIMULATED CONVENTIONAL AND BRANCHING TESTS

BRIEF

Requirement:

To compare a variety of computer-simulated conventional and branching tests and to extend the theoretical analysis of branching techniques.

Procedure:

Computer-simulated tests of two types were compared. One type consisted of conventional tests varying in length (5, 10, 15 items) and distribution of item difficulty indexes (all items at $p = .50$; normal, $p = .30 - .70$ and $p = .10 - .90$; and rectilinear, $p = .30 - .70$ and $p = .10 - .90$). The second consisted of branching tests varying in length (5, 10, 15 items to be answered by each examinee), number of items presented at each level of difficulty (1,2), and distribution of item difficulties comparable to those of the conventional tests. In addition, both types of test were varied in assumed item validity. The comparisons were made in terms of correlation between test scores and underlying ability (Lord's model).

Findings.

In tests with higher item validities ($r_{bis} = .60 - .90$), a branching test had a higher correlation with underlying ability than did any conventional test, for all three lengths studied.

Applicability:

This theoretical analysis supports an exploratory experimental study previously made. It indicates the research promise of tests with branching programs and provides useful guidelines for the design of further studies of programmed tests.

COMPARISON OF COMPUTER-SIMULATED CONVENTIONAL AND BRANCHING TESTS

CONTENTS

TABLES

FIGURES

COMPARISON OF COMPUTER-SIMULATED CONVENTIONAL AND BRANCHING TESTS

In line with its interest in unconventional testing techniques, the Behavior and Systems Research Laboratory (BESRL) has conducted several research studies of the branching techniques. Waters (1), in a theoretical study, found that a hypothetical five-item branching test correlated somewhat higher with underlying ability than did hypothetical five-item conventional tests. Bayroff and Seeley (2) obtained empirical indications that correlation between short branching tests and long conventional tests was considerably higher than the expected correlation between equally short conventional tests and long conventional tests. The present report continues the theoretical analysis and compares a variety of computer-simulated conventional and branching tests.

In the usual testing situation, each examinee takes all the items, and item sequence is the same for each examinee. It is possible, however, to have sequential or branching tests in which all examinees do not take the same items and the sequence of item presentation for an individual is some function of his performance on previous items; that is, an item answered correctly is followed by a more difficult item, an item answer incorrectly, by a less difficult item. The rationale for the latter procedure is that presentation of items based on an examinee's past performance allows each individual to take items that are progressively more appropriate to his own level of ability. It is conceivable that such a procedure would reduce testing time, and for a given amount of time would permit more accurate measurement of an individual's ability, principally by reducing opportunities for chance success by low ability examinees' attempting items too difficult for them.

TESTS

Conventional Tests

Five-, ten-, and fifteen-item hypothetical conventional (C) tests were evaluated. All tests were symmetric around $p = .50$, but varied in item difficulty distributions. The distributions investigated were all items at $p = .50$ (C50), roughly normal (CN), or rectilinear (CR). Each of the CN and CR tests was tried out with difficulty ranges of .30 through .70 and .10 through .90. Table 1 gives the C50, CN, and CR item difficulty distributions for the five-, ten-, and fifteen-item conventional tests.

Table 1

NUMBER OF ITEMS AT EACH P-VALUE FOR CONVENTIONAL TESTS

p-values	Five-Item Tests					Ten-Item Tests					Fifteen-Item Tests				
	all .50	Ranges .30-.70		.10-.90		all .50	Ranges .30-.70		.10-.90		all .50	Ranges .30-.70		.10-.90	
		N	R	N	R		N	R	N	R		N	R	N	R
.90				-	1				1	2				2	3
.80															
.70		-[a]	1	-	1		1	2	2	2		2	3	3	3
.60		-	1				2	2				3	3		
.50	5	-	1	-	1	10	4	2	4	2	15	5	3	5	3
.40		-	1				2	2				3	3		
.30		-	1	-	1		1	2	2	2		2	3	3	3
.20															
.10					1				1	2				2	3

[a]No five-item normally distributed tests were evaluated.

Branching Tests

One-Item-Per-Stage Tests. Six hypothetical one-item-per-stage (1-PS) branching tests were evaluated. The structure of these six tests is shown in Figures 1 through 6, together with the difficulty of each item. Two tests were studied at each of the three tests lengths (5, 10, and 15 items). One of the two tests covered a difficulty range of .30 through .70 and the other ranged from .10 through .90. The five-item-per-subject branching tests contained 15 items with each examinee responding to only five of the items. In the ten-item-per-subject tests, each examinee took ten of the 55 items in the test. The fifteen-item-per-subject tests were composed of 120 items. In each of the six tests, the first item (p = .50) was the same for all examinees, but the remaining items taken were determined by the examinee's performance on the immediately preceding item. If an examinee passed an item, he proceeded to a more difficult one; if he failed an item, he proceeded to an easier one. When the range of p-values in a test was .30 through .70, increases and decreases in difficulty between adjacent items were in steps of .05 for the five-item-per-subject test, .0222 for the ten-item-per-subject test, and .0143 for the fifteen-item-per-subject test. For the .10 through .90 range tests, the steps were .10 for the five-item-per-subject test, .0444 for the ten-item-per-subject test, and .0286 for the fifteen-item-per-subject test.

Two-Item-Per-Stage Tests. Four hypothetical two-item-per-stage (2-PS), ten-item-per-subject branching tests were evaluated. The structure of these four tests, and the distributions of item difficulties, is shown in Figure 7. Each of these tests was composed of 114 items. At each stage in these tests, the examinees took two items of the same difficulty level. The first two items taken by all examinees had p-values of .50. If the examinee passed both items in a pair, he branched to a more difficult item pair; if he passed one of the items in a pair, he branched to a pair of equal difficulty; if he failed both items in a pair, he proceeded to an easier pair of items. Items for two of the tests covered a difficulty range of .30 through .70, while the other two tests ranged from .10 through .90. For each of these difficulty ranges, one branching tests was developed by having equally spaced item pairs in the terminal row of the test (2-PS-E). The p-values of the item pairs in the other rows were determined from the terminal item pair values. For the other two-item-per-stage tests, 2-PS-U (one for each of the item difficulty ranges), the item pair p-values were determined by branching downward from the p = .50 item pair to the terminal row of item pairs. Using this procedure, the item pairs in the terminal rows were not equally spaced as in the 2-PS-E tests but were spaced so that the intervals between item pairs were smaller in the middle part of the difficulty ranges, and larger nearer the extreme difficulty values. Scores for all four of the two-item-per-stage tests ranged from 0 to 62.

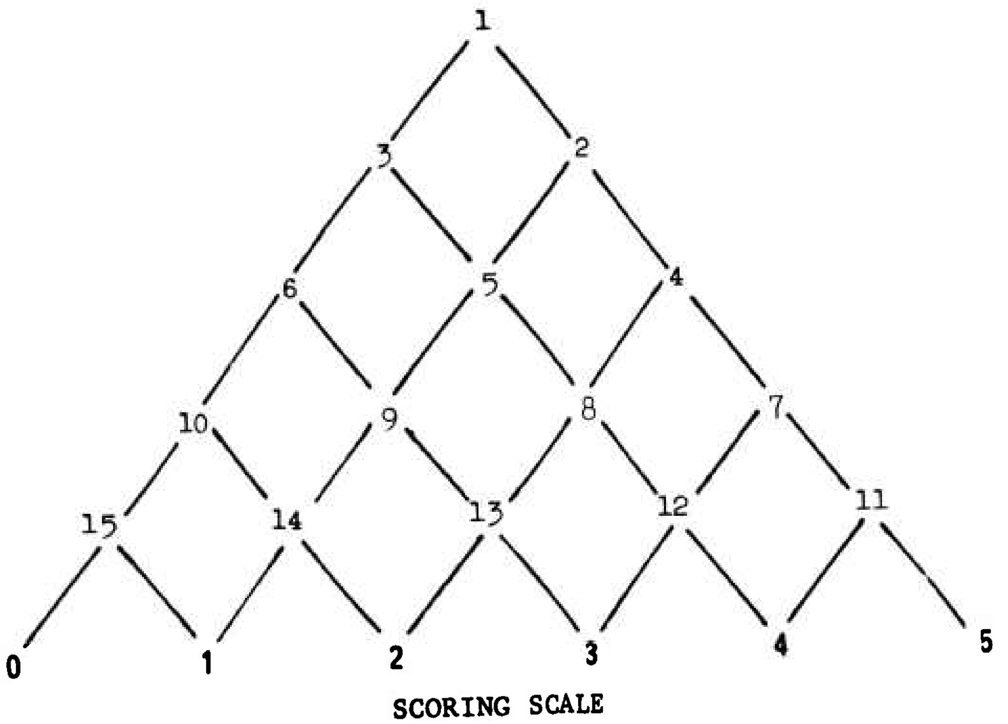

Figure 1. One-item-per-stage, five-item branching test with a difficulty range of .30-.70

DIFFICULTY VALUES FOR ITEMS SHOWN IN FIGURE 1

Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	9	.4500
2	.5500	10	.3500
3	.4500	11	.7000
4	.6000	12	.6000
5	.5000	13	.5000
6	.4000	14	.4000
7	.6500	15	.3000
8	.5500		

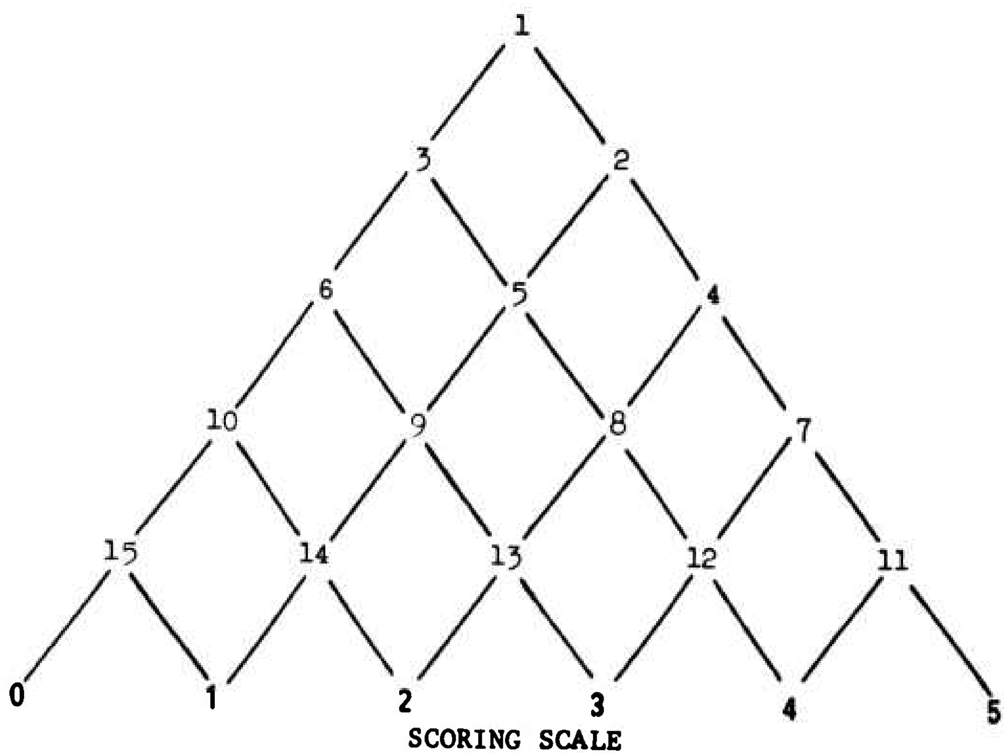

Figure 2. One-item-per-stage, five-item branching test with a difficulty range of .10-.90

DIFFICULTY VALUES FOR ITEMS SHOWN IN FIGURE 2			
Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	9	.4000
2	.6000	10	.2000
3	.4000	11	.9000
4	.7000	12	.7000
5	.5000	13	.5000
6	.3000	14	.3000
7	.8000	15	.1000
8	.6000		

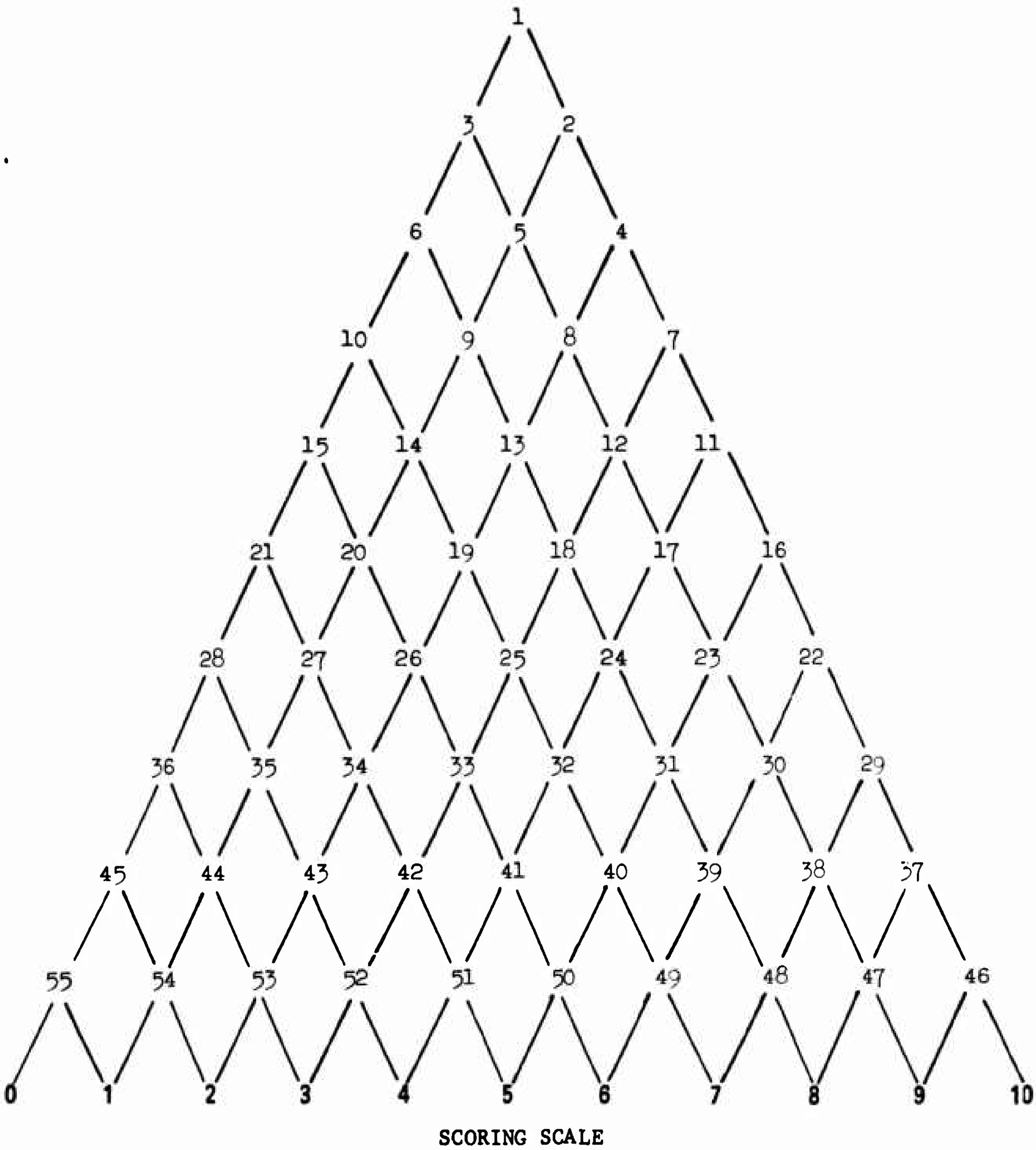

Figure 3. One-item-per-stage, ten-item branching test with a difficulty range of .30-.70 *(Difficulty values for items shown are given on page 7)*

DIFFICULTY VALUES FOR ONE-ITEM-PER-STAGE, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Figure 3)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	19	.4778	37	.6776
2	.5222	20	.4334	38	.6330
3	.4778	21	.3890	39	.5888
4	.5444	22	.6330	40	.5444
5	.5000	23	.5888	41	.5000
6	.4556	24	.5444	42	.4556
7	.5666	25	.5000	43	.4112
8	.5222	26	.4556	44	.3668
9	.4778	27	.4112	45	.3224
10	.4334	28	.3668	46	.6998
11	.5888	29	.6554	47	.6554
12	.5444	30	.6110	48	.6110
13	.5000	31	.5666	49	.5666
14	.4556	32	.5222	50	.5222
15	.4112	33	.4778	51	.4778
16	.6110	34	.4334	52	.4334
17	.5666	35	.3890	53	.3890
18	.5222	36	.3446	54	.3446
				55	.3002

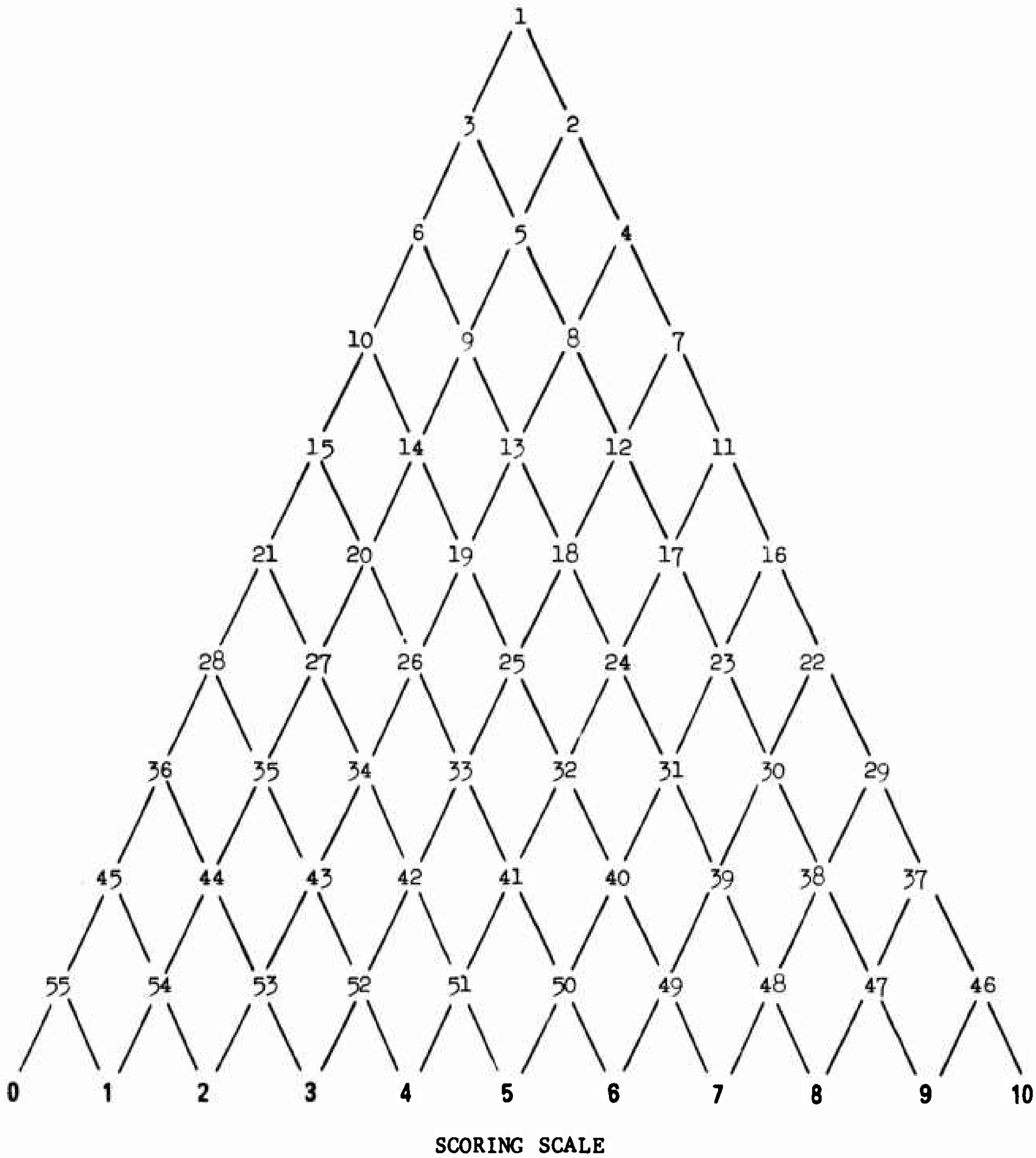

Figure 4. One-item-per-stage, ten-item branching test with a difficulty range of .10-.90
(Difficulty values for items shown are given on page 9)

DIFFICULTY VALUES FOR ONE-ITEM-PER-STAGE, TEN-ITEM BRANCHING TEST
WITH DIFFICULTY RANGE OF .10-.90 (Figure 4)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	19	.4556	37	.8552
2	.5444	20	.3668	38	.7664
3	.4556	21	.2780	39	.6776
4	.5888	22	.7664	40	.5888
5	.5000	23	.6776	41	.5000
6	.4112	24	.5888	42	.4112
7	.6332	25	.5000	43	.3224
8	.5444	26	.4112	44	.2336
9	.4556	27	.3224	45	.1448
10	.3668	28	.2336	46	.8996
11	.6776	29	.8108	47	.8108
12	.5888	30	.7220	48	.7220
13	.5000	31	.6332	49	.6332
14	.4112	32	.5444	50	.5444
15	.3224	33	.4556	51	.4556
16	.7220	34	.3668	52	.3668
17	.6332	35	.2780	53	.2780
18	.5444	36	.1892	54	.1892
				55	.1004

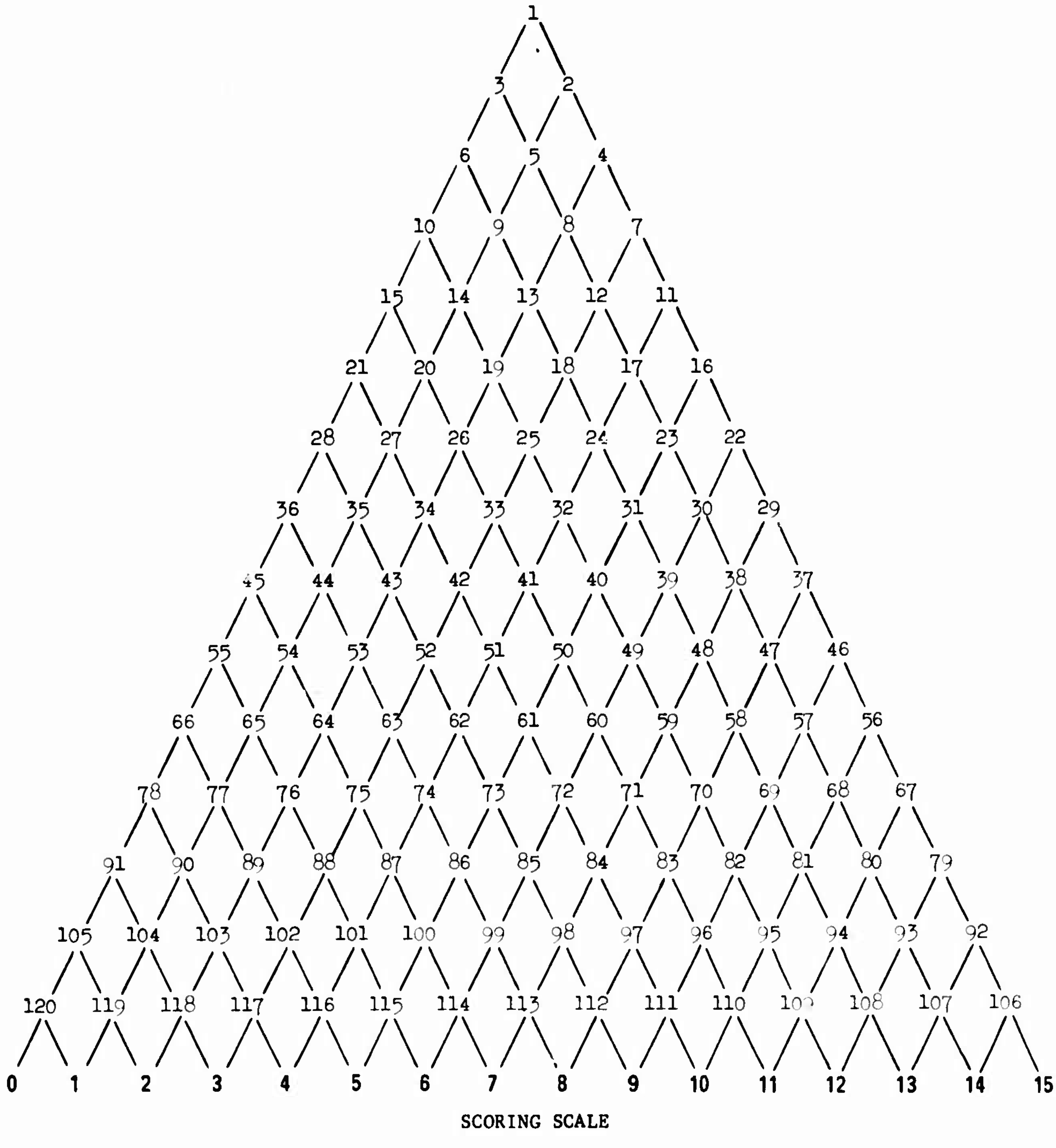

Figure 5. One-item-per-stage, fifteen-item branching test with a difficulty range of .30-.70
(Difficulty values for items shown are given on pages 11 and 12)

DIFFICULTY VALUES FOR ONE-ITEM-PER-STAGE, FIFTEEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Figure 5)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	21	.4285	41	.5000
2	.5143	22	.5858	42	.4714
3	.4857	23	.5572	43	.4428
4	.5286	24	.5286	44	.4142
5	.5000	25	.5000	45	.3856
6	.4714	26	.4714	46	.6287
7	.5429	27	.4428	47	.6001
8	.5143	28	.4142	48	.5715
9	.4857	29	.6001	49	.5429
10	.4571	30	.5715	50	.5143
11	.5572	31	.5429	51	.4857
12	.5286	32	.5143	52	.4571
13	.5000	33	.4857	53	.4285
14	.4714	34	.4571	54	.3999
15	.4428	35	.4285	55	.3713
16	.5715	36	.3999	56	.6430
17	.5429	37	.6144	57	.6144
18	.5143	38	.5858	58	.5858
19	.4857	39	.5572	59	.5572
20	.4571	40	.5286	60	.5286

DIFFICULTY VALUES FOR ONE-ITEM-PER-STAGE, FIFTEEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Continued)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
61	.5000	81	.6144	101	.4285
62	.4714	82	.5858	102	.3999
63	.4428	83	.5572	103	.3713
64	.4142	84	.5286	104	.3427
65	.3856	85	.5000	105	.3141
66	.3570	86	.4714	106	.7002
67	.6573	87	.4428	107	.6716
68	.6287	88	.4142	108	.6430
69	.6001	89	.3856	109	.6144
70	.5715	90	.3570	110	.5858
71	.5429	91	.3284	111	.5572
72	.5143	92	.6859	112	.5286
73	.4857	93	.6573	113	.5000
74	.4571	94	.6287	114	.4714
75	.4285	95	.6001	115	.4428
76	.3999	96	.5715	116	.4142
77	.3713	97	.5429	117	.3856
78	.3427	98	.5143	118	.3570
79	.6716	99	.4857	119	.3284
80	.6430	100	.4571	120	.2998

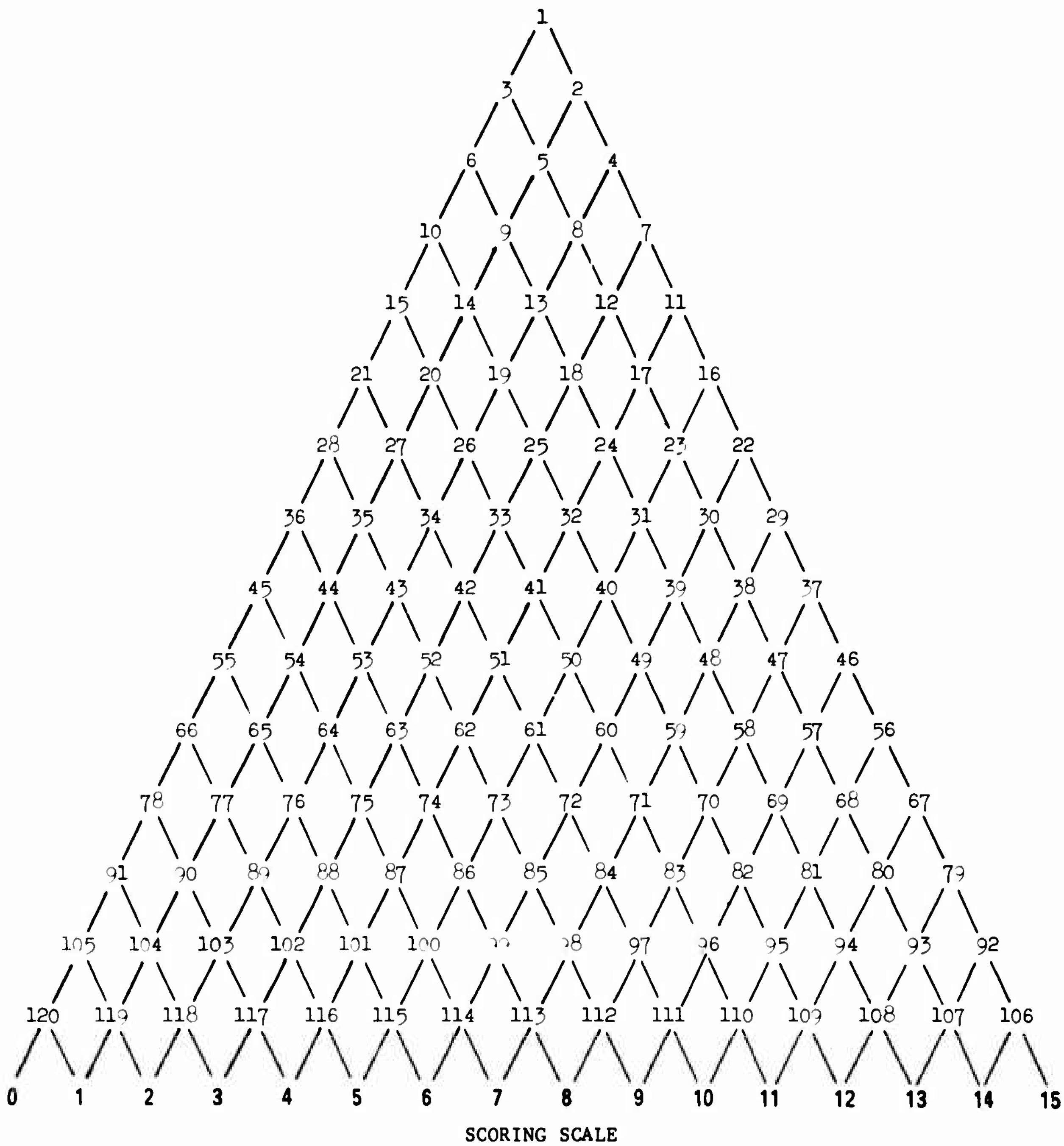

Figure 6. One-item-per-stage, fifteen-item branching test with a difficulty range of .10-.90
(Difficulty values for items shown are given on pages 14 and 15)

DIFFICULTY VALUES FOR ONE-ITEM-PER-STAGE, FIFTEEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .10-.90 (Figure 6)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	21	.3572	41	.5000
2	.5285	22	.6713	42	.4429
3	.4714	23	.6142	43	.3858
4	.5571	24	.5571	44	.3287
5	.5000	25	.5000	45	.2716
6	.4429	26	.4429	46	.7569
7	.5856	27	.3858	47	.6998
8	.5285	28	.3287	48	.6427
9	.4714	29	.6998	49	.5856
10	.4143	30	.6427	50	.5285
11	.6142	31	.5856	51	.4714
12	.5571	32	.5285	52	.4143
13	.5000	33	.4714	53	.3572
14	.4429	34	.4143	54	.3001
15	.3858	35	.3572	55	.2430
16	.6427	36	.3001	56	.7855
17	.5856	37	.7284	57	.7284
18	.5285	38	.6713	58	.6713
19	.4714	39	.6142	59	.6142
20	.4143	40	.5571	60	.5571

DIFFICULTY VALUES FOR ONE-ITEM-PER-STAGE, FIFTEEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .10-.90 (Continued)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
61	.5000	81	.7284	101	.3572
62	.4429	82	.6713	102	.3001
63	.3858	83	.6142	103	.2430
64	.3287	84	.5571	104	.1859
65	.2716	85	.5000	105	.1288
66	.2145	86	.4429	106	.8997
67	.8140	87	.3858	107	.8426
68	.7569	88	.3287	108	.7855
69	.6998	89	.2716	109	.7284
70	.6427	90	.2145	110	.6713
71	.5856	91	.1574	111	.6142
72	.5285	92	.8711	112	.5571
73	.4714	93	.8140	113	.5000
74	.4143	94	.7569	114	.4429
75	.3572	95	.6998	115	.3858
76	.3001	96	.6427	116	.3287
77	.2430	97	.5856	117	.2716
78	.1859	98	.5285	118	.2145
79	.8426	99	.4714	119	.1574
80	.7855	100	.4143	120	.1003

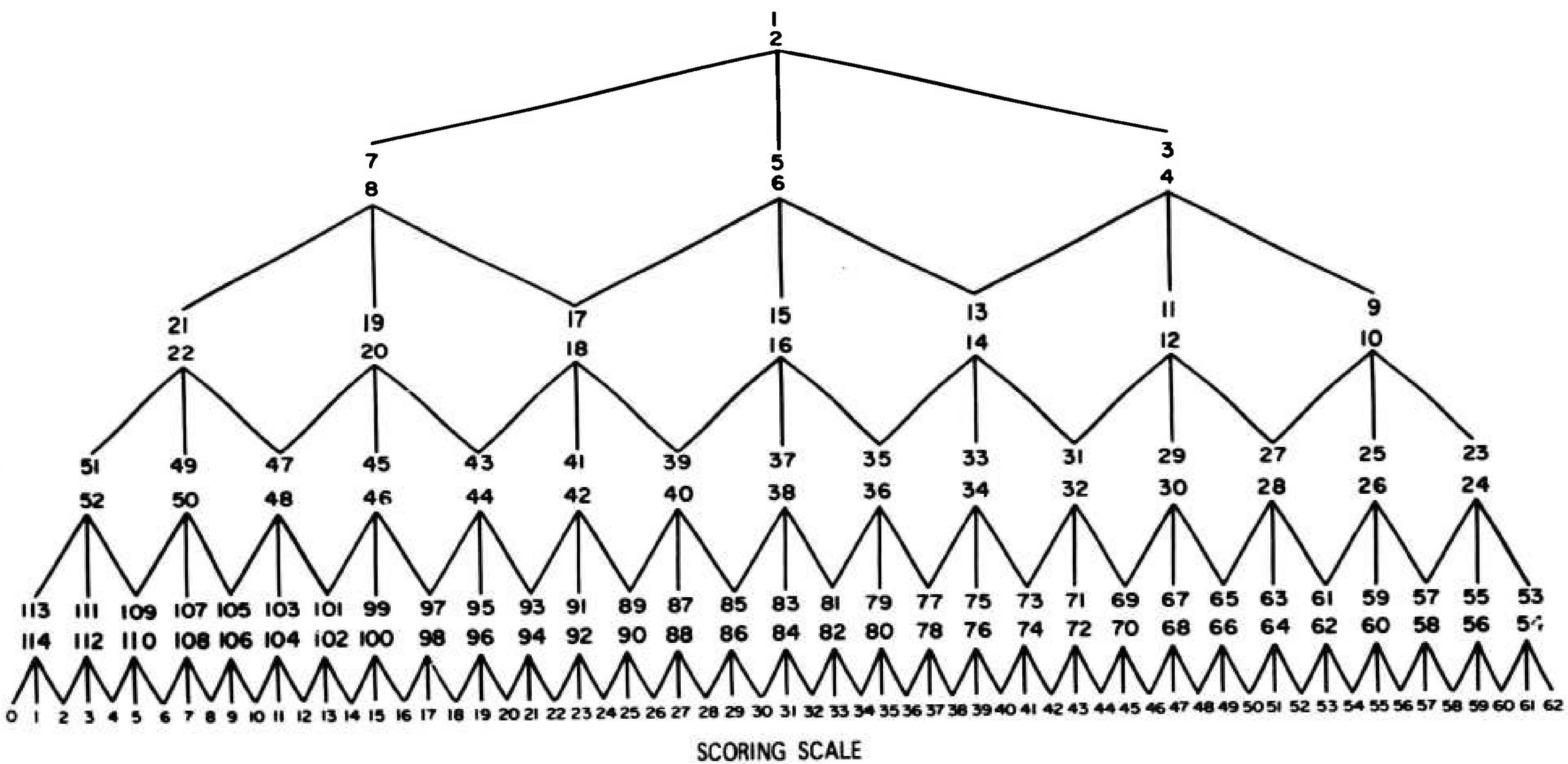

Figure 7. Diagram of two-item-per-stage, ten-item branching tests of varying difficulty distributions
(Distributions of item difficulties are given on pages 17 thru 24)

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-E, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Figure 7)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	20	.3933	39	.4733
2	.5000	21	.3400	40	.4733
3	.6067	22	.3400	41	.4467
4	.6067	23	.6867	42	.4467
5	.5000	24	.6867	43	.4200
6	.5000	25	.6600	44	.4200
7	.3933	26	.6600	45	.3933
8	.3933	27	.6333	46	.3933
9	.6600	28	.6333	47	.3667
10	.6600	29	.6067	48	.3667
11	.6067	30	.6067	49	.3400
12	.6067	31	.5800	50	.3400
13	.5533	32	.5800	51	.3133
14	.5533	33	.5533	52	.3133
15	.5000	34	.5533	53	.7000
16	.5000	35	.5267	54	.7000
17	.4467	36	.5267	55	.6867
18	.4467	37	.5000	56	.6867
19	.3933	38	.5000	57	.6733

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-E, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Continued)					
Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
58	.6733	77	.5400	96	.4200
59	.6600	78	.5400	97	.4067
60	.6600	79	.5267	98	.4067
61	.6467	80	.5267	99	.3933
62	.6467	81	.5133	100	.3933
63	.6333	82	.5133	101	.3800
64	.6333	83	.5000	102	.3800
65	.6200	84	.5000	103	.3667
66	.6200	85	.4867	104	.3667
67	.6067	86	.4867	105	.3533
68	.6067	87	.4733	106	.3533
69	.5933	88	.4733	107	.3400
70	.5933	89	.4600	108	.3400
71	.5800	90	.4600	109	.3267
72	.5800	91	.4467	110	.3267
73	.5667	92	.4467	111	.3133
74	.5667	93	.4333	112	.3133
75	.5533	94	.4333	113	.3000
76	.5533	95	.4200	114	.3000

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-U, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Figure 7)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	20	.4500	39	.4875
2	.5000	21	.4000	40	.4875
3	.5500	22	.4000	41	.4750
4	.5500	23	.6500	42	.4750
5	.5000	24	.6500	43	.4625
6	.5000	25	.6000	44	.4625
7	.4500	26	.6000	45	.4500
8	.4500	27	.5750	46	.4500
9	.6000	28	.5750	47	.4250
10	.6000	29	.5500	48	.4250
11	.5500	30	.5500	49	.4000
12	.5500	31	.5375	50	.4000
13	.5250	32	.5375	51	.3500
14	.5250	33	.5250	52	.3500
15	.5000	34	.5250	53	.7000
16	.5000	35	.5125	54	.7000
17	.4750	36	.5125	55	.6500
18	.4750	37	.5000	56	.6500
19	.4500	38	.5000	57	.6250

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-U, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .30-.70 (Continued)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
58	.6250	77	.5187	96	.4625
59	.6000	78	.5187	97	.4562
60	.6000	79	.5125	98	.4562
61	.5875	80	.5125	99	.4500
62	.5875	81	.5062	100	.4500
63	.5750	82	.5062	101	.4375
64	.5750	83	.5000	102	.4375
65	.5625	84	.5000	103	.4250
66	.5625	85	.4937	104	.4250
67	.5500	86	.4937	105	.4125
68	.5500	87	.4875	106	.4125
69	.5437	88	.4875	107	.4000
70	.5437	89	.4812	108	.4000
71	.5375	90	.4812	109	.3750
72	.5375	91	.4750	110	.3750
73	.5312	92	.4750	111	.3500
74	.5312	93	.4687	112	.3500
75	.5250	94	.4687	113	.3000
76	.5250	95	.4625	114	.3000

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-E, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .10-.90 (Figure 7)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	20	.2867	39	.4467
2	.5000	21	.1800	40	.4467
3	.7133	22	.1800	41	.3933
4	.7133	23	.8733	42	.3933
5	.5000	24	.8733	43	.3400
6	.5000	25	.8200	44	.3400
7	.2867	26	.8200	45	.2867
8	.2867	27	.7667	46	.2867
9	.8200	28	.7667	47	.2333
10	.8200	29	.7133	48	.2333
11	.7133	30	.7133	49	.1800
12	.7133	31	.6600	50	.1800
13	.6067	32	.6600	51	.1267
14	.6067	33	.6067	52	.1267
15	.5000	34	.6067	53	.9000
16	.5000	35	.5533	54	.9000
17	.3933	36	.5533	55	.8733
18	.3933	37	.5000	56	.8733
19	.2867	38	.5000	57	.8467

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-E, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .10-.90 (Continued)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
58	.8467	77	.5800	96	.3400
59	.8200	78	.5800	97	.3133
60	.8200	79	.5533	98	.3133
61	.7933	80	.5533	99	.2867
62	.7933	81	.5267	100	.2867
63	.7667	82	.5267	101	.2600
64	.7667	83	.5000	102	.2600
65	.7400	84	.5000	103	.2333
66	.7400	85	.4733	104	.2333
67	.7133	86	.4733	105	.2067
68	.7133	87	.4467	106	.2067
69	.6867	88	.4467	107	.1800
70	.6867	89	.4200	108	.1800
71	.6600	90	.4200	109	.1533
72	.6600	91	.3933	110	.1533
73	.6333	92	.3933	111	.1267
74	.6333	93	.3667	112	.1267
75	.6067	94	.3667	113	.1000
76	.6067	95	.3400	114	.1000

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-U, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .10-.90 (Figure 7)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
1	.5000	20	.4000	39	.4750
2	.5000	21	.3000	40	.4750
3	.6000	22	.3000	41	.4500
4	.6000	23	.8000	42	.4500
5	.5000	24	.8000	43	.4250
6	.5000	25	.7000	44	.4250
7	.4000	26	.7000	45	.4000
8	.4000	27	.6500	46	.4000
9	.7000	28	.6500	47	.3500
10	.7000	29	.6000	48	.3500
11	.6000	30	.6000	49	.3000
12	.6000	31	.5750	50	.3000
13	.5500	32	.5750	51	.2000
14	.5500	33	.5500	52	.2000
15	.5000	34	.5500	53	.9000
16	.5000	35	.5250	54	.9000
17	.4500	36	.5250	55	.8000
18	.4500	37	.5000	56	.8000
19	.4000	38	.5000	57	.7500

DIFFICULTY VALUES FOR TWO-ITEM-PER-STAGE-U, TEN-ITEM BRANCHING TEST WITH DIFFICULTY RANGE OF .10-.90 (Continued)

Item No.	Diff. Value	Item No.	Diff. Value	Item No.	Diff. Value
58	.7500	77	.5375	96	.4250
59	.7000	78	.5375	97	.4125
60	.7000	79	.5250	98	.4125
61	.6750	80	.5250	99	.4000
62	.6750	81	.5125	100	.4000
63	.6500	82	.5125	101	.3750
64	.6500	83	.5000	102	.3750
65	.6250	84	.5000	103	.3500
66	.6250	85	.4875	104	.3500
67	.6000	86	.4875	105	.3250
68	.6000	87	.4750	106	.3250
69	.5875	88	.4750	107	.3000
70	.5875	89	.4625	108	.3000
71	.5750	90	.4625	109	.2500
72	.5750	91	.4500	110	.2500
73	.5625	92	.4500	111	.2000
74	.5625	93	.4375	112	.2000
75	.5500	94	.4375	113	.1000
76	.5500	95	.4250	114	.1000

COMPUTATIONAL PROCEDURES AND ASSUMPTIONS

Statistical computations were based on a theoretical model presented by Lord (3). The model assumes that there is a trait or ability underlying the raw scores on a test, and that the probability of an examinee's responding correctly to a test item is a normal ogive function of his position on the ability dimension. Since item responses are a function only of scores on the ability continuum, they are independent of each other when ability is held constant. When all the items in a test are assumed to have the same biserial correlation (R_i) with ability, R_i^2 is an estimate of item intercorrelation. Three major steps are involved in obtaining the correlation between test score and underlying ability: The proportion of examinees passing each item is determined for each of the ability levels under consideration; the conditional distribution of test scores is obtained for each ability level; and the bivariate frequency distribution of test score and ability is obtained.

Proportion of Examinees at a Given Level of Ability Who Pass an Item

When the group tested is assumed to be normally distributed on ability, Lord's formulas (9) and (10) may be used to find the proportion of examinees who pass each of the test items when ability is held constant. In Lord's notation, a value of g_i (the z score corresponding to the p-value of item i at a specified ability level) is computed for each ability level under consideration by formula (9):

$$g_i = \frac{h_i - R_i \cdot c}{K_i}, \text{ where}$$

h_i = the z value corresponding to the population p-value of item i

R_i = the biserial correlation between item i and underlying ability

c = the z score representing the ability level being considered

$K_i = \sqrt{1 - R_i^2}$

Each g_i is converted to P_i (p-value of item i for examinees at a given ability level) by Lord's formula (10):

$$P_i = A\,(g_i) = \text{area of normal curve above the point } g_i.$$

These P_i values are computed for each ability level.

Conditional Test Score Distribution for Given Ability Levels

For conventional tests, the distribution of test scores at each of the specified ability levels may be computed by expansion of Lord's formula (11):

$$\prod_{i=1}^{n} (P_i + Q_i), \text{ where}$$

$\prod_{i=1}^{n}$ indicates the successive multiplication of the $(P_i + Q_i)$ terms

n = number of items in test

P_i = proportion passing item i for the given ability level

$Q_i = 1 - P_i$

Terms of this expansion give all possible ways of obtaining various test scores. Those terms which lead to the same test score are summed to obtain the distribution of test scores for a given ability level.

Although Lord does not discuss branching tests, his model is also applicable to this type of test. For a branching test, the proportion of examinees (at a specified ability level) following any given path may be determined by multiplying the P_i or Q_i values (as obtained by Lord's formulas 9 and 10) of the items which make up that path. If an item is passed, its P_i value is used; if an item is failed, its Q_i value is used. Such a proportion is computed for each path, and values for paths leading to the same test score are summed to obtain the test score distribution.

Bivariate Frequency Distribution of Test Score and Ability

For both conventional and branching tests, the bivariate distribution of test score and ability is obtained by multiplying the conditional test score distribution for each ability level by the ordinate value of the normal curve at that ability level (Lord's formula 14, applicable when a

normal distribution of ability is assumed). The test-ability correlation coefficient may be computed[1] from this scatterplot.

In the present study, the distribution of underlying ability in the theoretical sample of examinees was assumed to be normal with $\bar{X} = 0$ and $\sigma = 1.00$. Twenty-nine levels of ability, measured in standard scores ranging from +3.5 to -3.5 in steps of .25 were used. The biserial correlation between an item and ability was constant for all items in a given test. For each of the five- and ten-item-per-subject conventional and branching tests evaluated, the assumed biserial was varied from .30 to .90 in steps of .10. The fifteen-item tests were evaluated at biserials of .40, .60 and .80.

RESULTS AND DISCUSSION

Five-Item Tests

<u>Conventional Tests</u>. The correlation coefficients between test score and ability for the five-item conventional tests are shown in the first three rows of Table 2. For biserials of .30 through .70 (r_{ij} = .09 through .49), the all .50 (C50) test obtained the highest coefficients, and the .3[3] through .7 CR test yielded a higher relationship to the ability criterion than did the .1 through .9 CR test. At the .80 biserial (r_{ij} = .64), the .3 through .7 CR test yielded the highest coefficient, and the C50 test was next. Finally, at the assumed biserial of .90 (r_{ij} = .81), the wide range rectilinear test (.1 through .9 CR test) had the highest correlation coefficient, and the C50 test had the lowest coefficient of the three conventional tests. Overall, the C50 was best for low to moderate item intercorrelation; the moderate range (.3 through .7) and eventually the wider range (.1 through .9) tests were best for higher intercorrelations.

---

[1] A FORTRAN program which performs these computations was written for the GE 225 computer[2] by Mr. Sidney Sachs of the Computer Applications Branch, Behavior and Systems Research Laboratory. This program was used to obtain the test-ability coefficients reported. It should be noted that Brogden (4), Tucker (5), and Lord (3) have provided computationally easier formulas for obtaining the test-ability coefficients for conventional tests.

[2] The commercial designation is used only in the interest of specificity in reporting. Its use does not constitute indorsement by the Army or by BESRL.

[3] For simplicity in presentation, difficulty values mentioned in the text are hereafter expressed in one decimal place in contrast to the biserials which are in two decimal places.

Branching Tests. The results for the five-item-per-subject branching tests are shown in the last two rows of Table 2. The coefficient for the moderate range .3 through .7 test was higher than that for the wider range test for assumed biserials of .30 through .80; the .1 through .9 range test had the higher coefficient at r_{bis} = .90.

Comparison of Conventional and Branching Tests. One of the branching tests was superior to any of the conventional tests for $r_{bis} \geq .60$ ($r_{ij} \geq .36$). At the higher biserials, .70 through .90, both branching tests yielded higher coefficients than did any of the conventional tests. For the lower biserials, .30 through .50, the C50 conventional tests resulted in slightly higher coefficients than did either of the branching tests.

Table 2

TEST SCORE-ABILITY CORRELATION COEFFICIENTS FOR FIVE-ITEM-PER-SUBJECT CONVENTIONAL AND BRANCHING TESTS

Biserials	.30	.40	.50	.60	.70	.80	.90
Test			Correlation Coefficients[a]				
C (all .50)	482	601	696	769	823	858	871
C (.3 - .7, R)	473	591	686	762	819	861	887
C (.1 - .9, R)	43[illegible]	549	646	726	793	850	900
B (.3 - .7, 1-PS)	478	599	694	774	835	880	906
B (.1 - .9, 1-PS)	461	580	680	760	826	878	920

[a] Decimal points omitted.

Ten-Item Tests

Conventional Tests. The test score-ability correlation coefficients for the ten-item conventional tests are shown in the first five rows of Table 3. The C50 test had the highest coefficient for each biserial through .60. For these same biserials, all the .3 through .7 range tests were next highest and the .1 through .9 range tests were lowest. At r_{bis} = .70, the C50 and .30 through .7 tests were about equally effective, and yielded higher coefficients than the .1 through .9 tests. At biserials .80 and .90, the original situation was reversed and the C50 test had the lowest coefficients and the .1 through .9 tests the highest coefficients.

Branching Tests. The ten-item-per-subject branching test data are given in the bottom six rows of Table 3. The .3 through .7 1-PS tests tended to correlate higher with the criterion than did the .1 through .9 tests through a biserial of .60. Above this level the converse held. It should be noted that for all biserials, and any given item difficulty range, the 1-PS branching test correlated higher than any 2-PS test covering the same range. In fact, with only one exception (r_{bis} = .90), both the 1-PS .1 through .9 and .3 through .7 tests yielded higher coefficients than did any of the 2-PS tests. The 2-PS-E tests correlated higher with the ability criterion than did the 2-PS-U tests.

Table 3

TEST SCORE-ABILITY CORRELATION COEFFICIENTS FOR TEN-ITEM-PER-SUBJECT CONVENTIONAL AND BRANCHING TESTS

Biserials	.30	.40	.50	.60	.70	.80	.90
Test			Correlation Coefficients[a]				
C (all .5)	614	728	807	850	891	905	898
C (.3 - .7, N)	608	723	802	856	890	909	910
C (.3 - .7, R)	604	719	799	854	890	911	917
C (.1 - .9, N)	586	702	786	844	886	913	929
C (.1 - .9, R)	583	680	767	830	877	913	941
B (.3 - .7, 1-PS)	612	728	808	866	904	926	931
B (.3 - .7, 2-PS-E)	520	642	737	809	863	898	915
B (.3 - .7, 2-PS-U)	512	633	721	799	851	885	898
B (.1 - .9, 1-PS)	601	719	801	862	905	934	953
B (.1 - .9, 2-PS-E)	531	655	751	825	881	921	948
B (.1 - .9, 2-PS-U)	519	640	729	808	862	899	918

[a]Decimal points omitted.

Comparison of Conventional and Branching Tests. One of the 1-PS branching tests was superior to any of the conventional tests for biserials above .40 (the .3 through .7 1-PS was highest at r_{bis} = .50 and .60; the .1 through .9 1-PS was highest at r_{bis} = .70 through .90). At a biserial of .30, the C50 test coefficient was slightly higher and

at r_{bis} = .40 the C50 conventional and .3 through .7 1-PS branching tests had the largest coefficients. The 2-PS branching tests compared favorably with the best conventional test only at very high biserials.

Fifteen-Item Tests

Conventional Tests. All fifteen-item tests were evaluated at biserials of .40, .60, and .80. The test score-ability correlation coefficients for the five conventional tests are given in the first five rows of Table 4. The C50 test had the highest coefficient at biserials of .40 and .60. At a biserial of .80, the .1 through .9 tests (both N and R) did best. A comparison of the .3 through .7 and .1 through .9 tests across the three biserials showed that the narrower range tests received higher coefficients at the lower biserials (.40 and .60) and the wider range tests did better for the high biserial (.80). This general trend was consistent with the results obtained for the five- and ten-item conventional tests. For tests of a given range of item difficulties, those with approximately normally distributed item difficulties were superior to those with rectilinear difficulty distributions at the .40 biserial and did less well than their rectilinear counterparts at a biserial of .80. At r_{bis} = .60, no difference was obtained between the .3 through .7 N and R tests, but the .1 through .9 N test was superior to the R test of the same range. This same trend was also found for the ten-item conventional tests. In general, as biserials (and thus item intercorrelations) increased, wider range tests and tests with more rectilinear item difficulty distributions did progressively better.

Branching Tests. Data for the two fifteen-item branching tests are given in the last two rows of Table 4. The .3 through .7 test correlated higher with the ability criterion at the .40 biserial, while the .1 through .9 test yielded the highest coefficient at r_{bis} = .80. The two branching tests were essentially equivalent at the .60 biserial.

Comparison of Conventional and Branching Tests. Both branching tests yielded higher coefficients than did any of the conventional tests for biserials of .60 and .80. At the .40 biserial, the C50 test was essentially equivalent to the .3 through .7 branching test.

Effects of Test Length

Table 5 gives the increments in test score-ability coefficients as the tests were increased in length from five to fifteen items. Increasing the number of items from five to ten resulted in increments in correlation about twice as large as those obtained by increasing test length from ten to fifteen items. Increases in test length led to higher test score-ability coefficients for the lower biserial values. There appeared to be little difference between conventional and branching tests in terms of the effects of increasing test length.

Table 4

TEST SCORE-ABILITY CORRELATION COEFFICIENTS FOR FIFTEEN-ITEM-PER-SUBJECT CONVENTIONAL AND BRANCHING TESTS

Biserials	.40	.60	.80
Test		Correlation Coefficients[a]	
C (all .50)	792	896	923
C (.3 - .7, N)	787	894	928
C (.3 - .7, R)	785	894	930
C (.1 - .9, N)	764	884	936
C (.1 - .9, R)	751	877	937
B (.3 - .7, 1-PS)	793	903	943
B (.1 - .9, 1-PS)	786	902	953

[a] Decimal points omitted.

Table 5

INCREMENTS IN TEST SCORE-ABILITY CORRELATION COEFFICIENTS WITH INCREASE IN TEST LENGTH

Biserials		.30	.40	.50	.60	.70	.80	.90
					Correlation Coefficients[a]			
C (all .50)	5-10	132	127	111	090	068	047	027
	10-15		064		037		018	
C (.3 - .7, N)	5-10							
	10-15		064		038		019	
C (.1 - .9, N)	5-10							
	10-15		062		040		023	
C (.3 - .7, R)	5-10	131	128	113	092	071	050	030
	10-15		066		040		019	
C (.1 - .9, R)	5-10	129	131	121	104	084	063	041
	10-15		071		047		024	
B (.3 - .7, 1-PS)	5-10	134	129	114	092	069	046	025
	10-15		065		037		017	
B (.1 - .9, 1-PS)	5-10	137	133	115	094	071	048	025
	10-15		066		039		018	

[a] Decimal points omitted.

OVERVIEW

Both conventional and branching test data showed that tests with the least spread of item difficul ies yielded the highest correlation coefficients with underlying abilit, when low to moderate item biserials were assumed. For medium to high biserials, the moderate range and wide range tests tended to yield coefficients of about the same magnitude. The wide range tests generally did best when very high biserials were assumed. These data are consistent with the 9- and 18-item test data reported by Brogden (4). The shift in the relative effectiveness of the narrower and wider range tests tended to take place earlier when test length was increased.

For the lowest biserial assumed (.30), the C50 test was the only conventional test which correlated higher with the ability criterion than did the best branching test. At biserials of .40 and .50, the ten- and fifteen-item branching tests covering a .3 to .7 range and the C50 test were essentially equivalent. For biserials of .60 and above, one of the branching tests always did better than anv of the conventional tests. In general, the differences in correlation with underlying ability were small but systematic. Since the data were by definition errorless, greater significance may be attached to these differences than would be the case with empirical data.

A comparison of one-item-per-stage and two-item-per-stage branching tests (at the ten-item test length) indicated that the one-item-per-stage tests had uniformly higher coefficients than did the two-item-per-stage tests of the same range. In view of these results, it would not seem profitable to use the more complex two-item-per-stage structure in the development of branching tests for the purpose of maximizing overall correlation.

LITERATURE CITED[a]

1. Waters, Carrie J. Preliminary evaluation of simulated branching tests. BESRL Technical Research Note 140. June 1964. (AD 602 978)

2. Bayroff, A. G. and L. C. Seeley. An exploratory study of branching tests. BESRL Technical Research Note 188. June 1967. (AD 655 263)

3. Lord, F. M. A theory of test scores. Psychometric Monograph No. 7, 1952.

4. Brogden, H. E. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. Psychometrika, 1946, 11, 197-213.

5. Tucker, L. R. Maximum validity of a test with equivalent items. Psychometrika, 1946, 11, 1-3.

ADDITIONAL REFERENCES

Bayroff, A. G., J. A. Thomas, and A. A. Anderson. Construction of an experimental sequential item test. BESRL Research Memorandum 60-1. January 1960.

Krathwohl, D. R. and R. J. Huyser. The sequential item test (SIT). American Psychologist, 11, 1956, p. 419.

Seeley, L. C., Mary A. Morton, and A. A. Anderson. Exploratory study of a sequential item test. BESRL Technical Research Note 129. December 1962. (AD 411 225)

[a] BESRL publications available from the U. S. Defense Documentation Center (DDC), formerly ASTIA, are indicated by inclusion of the assigned AD number in parentheses at the end of the citation. For publications not accessioned by DDC, the publishing agency or source is given.

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Behavior and Systems Research Laboratory, OCRD, Arlington, Virginia	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

COMPARISON OF COMPUTER-SIMULATED CONVENTIONAL AND BRANCHING TESTS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

Carrie W. Waters

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1970	45	5

8a. CONTRACT OR GRANT NO	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DA R&D Proj. No. 2Q024701A721	Technical Research Note 216
c. Enlisted Manpower	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office, Chief of Research and Development, DA

13. ABSTRACT

In the conduct of continuing research to improve screening of potential enlisted manpower, the Enlisted Manpower Work Unit has explored the potential contribution of tests with branching programs. Such tests permit greater individualization of testing than do conventional tests by testing with items appropriate to each examinee's ability level. The present publication reports on a comparison of computer-simulated branching and conventional tests which varied in length, distribution of item difficulty, and item validity. Comparison was in terms of correlation between underlying ability and test scores (Lord's model). The model assumes that there is a trait or ability underlying the raw scores on a test and that the probability of examinee's responding correctly to a test item is a normal ogive function of his position on the ability dimension.

The principal finding was that in tests with higher item validities (r_{bis} = .60 - .90), a branching test had higher correlation with underlying ability than did any of the conventional tests, for all three lengths studied.

This theoretical analysis supported an exploratory experimental study previously made and provides useful guidelines for further research with branching tests.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Automated testing						
Screening and classification tests						
Programmed testing						
Computer-simulation						
Conventional tests						
Branching tests						
Branching techniques						
Theoretical analysis						
Psychological measurement						
Psychometrics						
Military psychology						
Test theory						